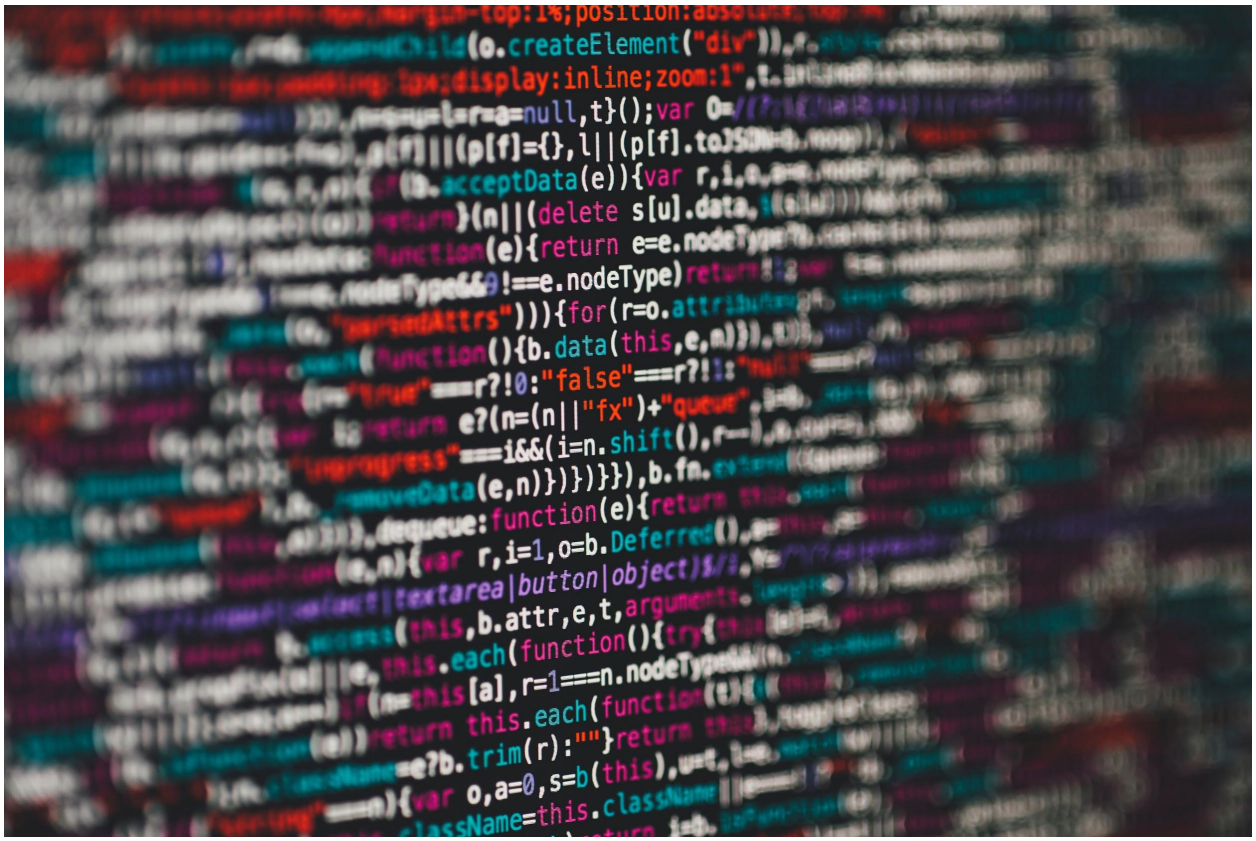
Citizen Science Projects (MOOC) 4.5

Woods, Mel; Coulson, Saskia; Ajates, Raquel; Amditis, Angelos ; Cobley, Andy; Domian, Dahlia

Publication date:
2020



Historically, the evolution of language has given people the ability to share information and knowledge. Technology has changed the way information spreads. When people invented printing and books, we had a tool for sharing knowledge that was costly in terms of writing and production, but this method could attribute the knowledge to its author. The information was distributed based on the number of printed copies made.

Now that we have the internet, data can be duplicated freely and can reach every corner of the world in an instant. It is easy to separate the author from the work, so, at first, authors were reluctant to massively distribute their data in this uncontrolled way. Today, though, we have the open data movement which propels us towards an internet where data are free to use, reusable and easy to distribute, but can also easily be attributed and licensed. Linked open data are part of this movement. In exchange for access to open data, we now link to their source on the web.

###Open data sources

In citizen science, it is particularly important to offer open access to the data collected by the users. It is good practice to give citizen scientists something back from their participation in the project. Having the data come back to them enriched and integrated with other sources in an easy-to-use format is a valuable initial reward. Some projects go beyond that, and open data

becomes the mission of the project. [OpenStreetMap](https://openstreetmap.org) is a perfect example of a citizen science project that was created to ensure the existence of open cartography. The foundation's mission statement says:

> "The OpenStreetMap Foundation is there to protect the OSM data to keep it free and open."

Many citizen science data platforms such as [iNaturalist](https://www.inaturalist.org) or [Spotteron](https://www.spotteron.net) offer an export to CSV function for occasional downloads, as well as open APIs to create a permanent connection to the data.

##Benefits and challenges

The government has embraced the open data movement. For them, open data facilitates transparency, accountability and public participation. It empowers citizens and strengthens democracy, and at the same time supports technological innovation and economic growth by enabling third parties to develop services.

For scientists, open data allows for global studies aiming to understand global problems such as climate change and helps to solve problems such as disease, crime or famine. Computer engineers also benefit because open data provides input for machine learning training.

In some cases, an open data approach is not appropriate. For example, projects working with indigenous communities that make the data public without their permission could result in resource extraction and abrogation of rights. For more details, see the [Indigenous Data Sovereignty
networks](https://www.rd-alliance.org/groups/international-indigenous-data-sovereignty-ig) for more details.

##Who is doing this well?

Tim Berners-Lee, the inventor of the World Wide Web and Linked Data initiator, suggested a 5-star measurement system to assess data openness. We can see these stars as steps towards a data sharing best practice. These steps are:

1. Publish the data under an open licence.
2. Make data available in a structured format (e.g. a table).
3. Use an open format (e.g. a CSV file).
4. Provide the data as a queryable web service in a way that others can extract part of the data using a URI (a 'uniform resource identifier' - these are used to identify a resource on the web, similar to a 'uniform resource locator (URL)' that locates that resource on the internet).
5. Provide metadata and connect to semantics to integrate the necessary context information so that users can correctly interpret the data and combine them with other similar data.

##Explore open data and share interesting datasets

A recent report from the Joint Research Centre called ["An inventory of citizen science activities for environmental policies"](http://data.europa.eu/euodp/data/dataset/jrc-citsci-10004) identified 503 projects in the environment field. As the number of citizen science projects increases, the need for standard practices has become even more critical. The [Open Geospatial Consortium (OGC)](https://www.opengeospatial.org/) and the WeObserve project are working on developing best practices on how to apply existing OGC standards for citizen science.

##Coordinating projects to achieve a bigger impact

The increasing number of citizen science initiatives have led to the creation of a fragmented landscape of data repositories. Projects have their resources available under different models, standards and technologies. In an effort to deter silos of resources, the [Global Earth Observation System of Systems (GEOSS)](https://www.earthobservations.org/geoss.php) offers a single access point to [Earth Observation data](https://www.geoportal.org). This portal connects users to various environmental monitoring systems around the world, while promoting the use of common technical standards. Similarly, the [European Open Science Cloud](https://www.eosc-portal.eu) (EOSC) provides a virtual environment with open and seamless services for storage, management, analysis and re-use of research data. Integrating available resources with mechanisms like these and leveraging their data management principles gives us a simple way to deal with open data obligations and to access research data across different disciplines while also promoting the use of open solutions and common standards for data sharing.

##Share your experience

+ Have you contributed to or accessed open data?
+ What were the data about, and how did you use them?